John von Neumann Institute for Computing

Sequence Specific DNA Recognition by EcoRV

M. Zahran, P. Imhof, J. C. Smith

published in

From Computational Biophysics to Systems Biology (CBSB08),
Proceedings of the NIC Workshop 2008,
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,
Walter Nadler, Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **40**, ISBN 978-3-9810843-6-8, pp. 417-420, 2008.

http://www.fz-juelich.de/nic-series/volume40

Sequence Specific DNA Recognition by EcoRV

Mai Zahran[1], **Petra Imhof**[1], **and Jeremy C. Smith**[1,2]

[1] Computational Molecular Biophysics, IWR, University of Heidelberg, Heidelberg, Germany
E-mail: {mai.zahran, petra.imhof}@iwr.uni-heidelberg.de

[2] Center for Molecular Biophysics, Oak Ridge National Laboratory,
P.O. Box 2008 MS6164 Oak Ridge, TN 37831-6164, USA
E-mail: smithjc@ornl.gov

EcoRV is a restriction enzyme produced by Escherichia Coli that destroys invading DNA by cleaving it at a GATATC sequence, as a defence mechanism against viral attacks. EcoRV sharply bends its specific DNA site, by approximately 50°, directly at the center TA step, helping to facilitate correct orientation of the scissile phosphate, the catalytic site of the enzyme and divalent metal ions. We are investigating the physical origins of sequence specificity in EcoRV endonuclease. Computer simulations are performed of three free DNA sequences in aqueous solution, starting from a B-form, in order to understand to what extent the bending is intrinsic to the DNA itself. The results contribute to understanding why the cognate sequence is recognized specifically by the EcoRV restriction enzyme.

1 Introduction

One of the central issues of modern molecular biology and biophysics is to understand the interactions stabilizing complexes in solution, and particularly how a small sequence change can lead to a significant difference in affinity. Protein/DNA interactions provide an important class of biomolecular complexes for studying such issues. Molecular dynamics (MD) simulation techniques, with which individual interactions and dynamics at the atomic level can be probed explicitly, provide a means complementary to experimental techniques, of determining details of molecular structure and interaction.[8] Restriction endonucleases provide interesting model systems for the investigation of sequence-specific protein/DNA interactions. The ability of bacterial cells to resist invading foreign DNA is wholly dependent upon the extraordinarily high fidelity of this recognition process, in which target sites are selected from an enormous molar excess of structurally similar non-specific DNA.[1,3] We are investigating the origins of sequence specificity in EcoRV endonuclease in order to elucidate how molecular interactions and induced fit operate to ensure selectivity for blunt-ended cleavage at the center step of GATATC. EcoRV sharply bends its specific DNA site by approximately 50°, directly at the center TA step, helping to facilitate proper juxtaposition of the scissile phosphate, the catalytic site of the enzyme and divalent metal ions.[7] To examine the origins of EcoRV sequence discrimination, a detailed kinetic and crystallographic study has been performed of the interaction of EcoRV with the cognate sequence GATATC (TA), and two non-cognate sequence, GAATTC (AT) which is the cognate sequence recognized by an other restriction enzyme, EcoRI, and it's mutant, GAAUTC (AU).[2] Examination of DNA binding and bending by equilibrium and stopped-flow fluorescence quenching and fluorescence resonance energy transfer (FRET) methods demonstrate that the capacity of EcoRV to bend the AT sequence site is severely limited, and full bending of AU sequence is achieved at only a threefold reduced rate compared

with the cognate complex (TA). The above results demonstrate that the DNA sequences, although swapping only 2 nucleotides, induce a change of the bending-cleavage mechanism. This change does not imply large conformation adaptations of the protein rather small perturbations leading to the rearrangement of the divalent metal ion binding sites. In the present work, the structures of the free DNA sequences in water are probed. We investigate how the swap of the center base pair leads to a change of the behaviour of the free DNA in water. For this purpose, we performed MD simulations of the three DNA sequences; TA, AT and AU free in aqueous solution.

2 Materials and Methods

Three 14-bp B-DNA molecules were examined with different central nucleotides but the same flanking sequences and same nucleotide content (**TA**-sequence: 5'-dAGAAGA**TA**TCTTGA-3', **AT**sequence: 5'-dAGAAGA**AT**TCTTGA-3', **AU**-sequence: 5'-dAGAAGA**AU**TCTTGA-3'). Standard B-DNA starting structures were generated using the program NAB[6]. The setup of the system was performed with the program CHARMM[9] and the charmm27 force field[10]. Each system was neutralized by adding 28 Na^+ counterions, and an excess of Na^+ and Cl^- ions were added, corresponding to a physiological concentration of 150mN NaCl. The system was solvated with a buffer of explicit water extending of 15 Å in each direction in a cubic box (x=70Å, y=70Å, z=70 Å). Minimizations, heating, equilibration and production runs were performed with the program NAMD[11]. For each of the three systems 20ns MD was performed in a NPT ensemble at 1-atm pressure and 300K. Structural analysis and calculation of the free energy were performed using standard programs; 3DNA[5], Gromacs[4] tools and home made scripts.

3 Results and Discussion

Differences of two nucleotides at the center steps have been shown to be sufficient to hinder full bending and thus cleavage of the DNA in complex with EcoRV restriction endonuclease.[2] We examine here whether or not differences can be seen between these three free DNA molecules in water. We analyzed the molecular origin of bending in terms of local helical parameters of the three DNA structures at the center step calculated with the program 3DNA[5]. The parameters monitored include the local roll angle and the local tilt at the center step. The roll angle measures rotation of a base pair plane about its long axis. This motion creates an angle, narrowing toward the major groove for positive roll, between two otherwise parallel adjacent base pairs. Tilt arises from rotation about a base pair short axis and is generally less than half as large as roll. We calculated the free energy profile at the center step of each DNA sequence as a function of there two parameters. Table 1 shows some local helical parameters for the DNA molecules crystallized in complex with the protein EcoRV, and in the DNA during the MD simulation. The three DNA molecules appear to have different properties for the roll angle at the center step when it is complexed to EcoRV or when it is free in aqueous solution. Indeed the cognate free TA molecule has a minimum free energy for an angle of $10^\circ \pm 0.98^\circ$, whereas the free AT molecule has an angle of $5^\circ \pm 0.65^\circ$ and the free AU molecule is about $3^\circ \pm 0.74^\circ$. Within 5 kJ/mol of the minimum free energy the roll angle for the TA sequence ranges from -10 to +25°,

whereas that for AT reaches only -5 to +15°. The AU sequence explores a wider range than AT but less than the TA sequence, varying between -10 to +20° within 5kJ/mol of the minimum. In contrast to the roll angle, the tilt angle behaves in a similar way for the 3 DNA molecules. The atomic fluctuations (Figure 1) show that the 3 DNA molecules fluctuate in

Parameter	TA[a]	AT[a]	AU[a]	TA[b]	AT[b]	AU[b]
Roll (deg.)	49.9	28.8	53.2	10 ±0.98	5 ±0.65	3 ±0.74
Tilt (deg.)	-3.24	0.87	4.01	0.33 ±0.08	0.63 ±0.05	1.64 ±0.09

[a] DNA crystallographic structures. The respective pdb codes are 1SX8, 2BOD, 2BOE. The values are taken from Hiller[2].
[b] B-DNA structures simulated in water. The values correspond to the angle with the minimum free energy calculated from the 20ns MD simulation.

Table 1. Conformational diffusion constants and corresponding relaxation times.

the same regions, mainly at the 2 termini (atoms 1 to 40; 400 to 500 and 850 to 890 around 0.6 nm for TA sequence, 0.3 nm for AT sequence and 0.8 for AU sequence). The rest of the atoms fluctuate around 0.15 nm, and a periodicity can be observed. The higher values correspond to the fluctuation of the atoms from the backbone (sugar and phosphate), and the lower values to the base pairs. The three structures behave mostly in a very similar

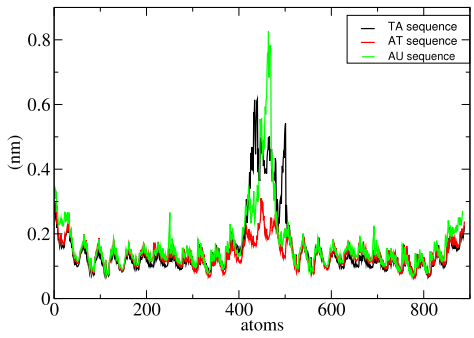

Figure 1. Atomic fluctuation of the 3 DNA sequences during the MD simulation. The first strand is from the atom 1 to 448; the second is from 449 to 891.

way. However the roll angle at the center step shows a none negligible difference.

4 Conclusion

The three DNA molecules simulated here differ by only 1 or 2 nucleotides at the central step. This small difference is sufficient for the EcoRV restriction endonuclease to recognize

only one sequence specifically. The simulations show that the atomic fluctuations and some local helical parameters are very similar for the three sequences, even at the central step. However, the TA sequence appears to be the most flexible of the three, as can be seen by the broader range of roll angles sampled as compared to the other sequences during the 20 ns of the simulations. This finding suggests that recognition of the sequence specifically by the EcoRV restriction enzyme is due in part to inherent structural tendencies of the cognate DNA sequence.

References

1. Kovall, R. A. and Matthews, B. W. *Type II restriction endonucleases: structural, functional and evolutionary relationships*, Curr. Opin. Chem. Biol. **3**, 578-583, 1999.
2. D. A. Hiller, A. M. Rodriguez and J. J. Perona *Non-cognate enzyme-DNA complex: structural and kinetic analysis of EcoRV endonuclease bound to the EcoRI recognition site GAATTC*, J. Mol. Biol. **354**, 121-136, 2005.
3. D. Parry, S. A. Moon, H. H. Liu,P. Heslop, B. A. Connolly *3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures*, J. Mol. Biol. **331**, 1005-1016, 2003.
4. E. Lindahl, B. Hess and D. van der Spoel *GROMACS 3.0: A package for molecular simulation and trajectory analysis*, J. Mol. Mod. **7**, 306-317, 2001.
5. Xiang-Jun Lu and Wilma K. Olson *Non-cognate enzyme-DNA complex: structural and kinetic analysis of EcoRV endonuclease bound to the EcoRI recognition site GAATTC*, J. Mol. Biol. **31**, 5108-5121, 2003.
6. T. Macke and D.A. Case *Modeling unusual nucleic acid structures. In: Leontes NB, SantaLucia J Jr. , editors.*, In Molecular Modeling of Nucleic Acids.Washington, DC: American Chemical Society, 379-393, 1998.
7. Winkler FK, Banner DW, Oefner C, Tsernoglou D, Brown RS, Heathman SP, Bryan RK, Martin PD, Petratos K, Wilson KS. *The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments.*, EMBO J. **12**, 1781-1795, 1993.
8. Sen S, Nilsson L. *Structure, interaction, dynamics and solvent effects on the DNA-EcoRI complex in aqueous solution from molecular dynamics simulation.*, Biophys J. **77**, 1782-1800, 1999.
9. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations.*, Journal of Computational Chemistry **4**, 187-217, 1983.
10. MacKerell, Jr. AD, Banavali N, Foloppe N *Development and current status of the CHARMM force field for nucleic acids*, Biopolymers **56**, 257-265, 2001.
11. James C. Phillips, Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kale, and Klaus Schulten. *Scalable molecular dynamics with NAMD.*, Journal of Computational Chemistry **26**, 1781-1802, 2005.